An Overview of Search Tools

Most online "searching tools" can be categorized as either a **subject tree, search engine, or virtual library**. In order to be successful in your searches, it is important to always know which kind of online resource you are using. This page provides a brief overview to explain the differences between these classes of online tools.

My other search pages: <u>Methodology</u> , <u>search tools</u> , <u>Advanced Features</u> , <u>Information Space</u> , <u>Specialized Tools</u>.

Types of online Search tools:
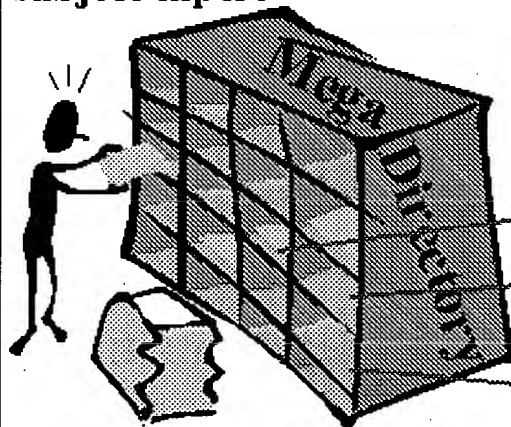
- <u>Subject Trees</u> (Yahoo) - Manually built "index" or card catalog
- <u>Search Engines</u> (Alta-Vista) - Very Large databases
- <u>Virtual Libraries</u> (Joe's guide to widgets) - Built by subject experts
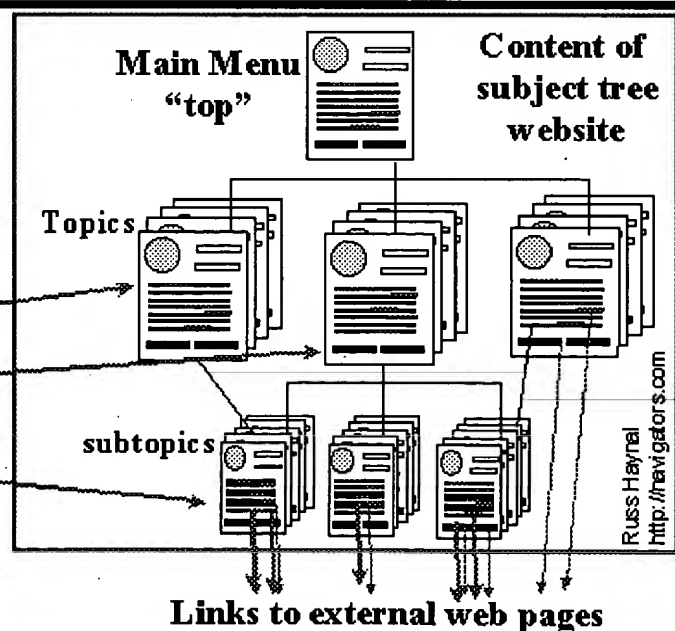

Each Tool has Strengths and Weaknesses

Pick the right tool...

Subject Trees:

Filer may not be a subject expert

URL's & Descriptions submitted by users

Main Menu "top"

Content of subject tree website

Topics

subtopics

Links to external web pages

Russ Haynal
http://navigators.com

Subject Tree - Key Features

- Main menu presents a list of menu choices
- Web pages are created manually
- Web sites must "announce" or "register" themselves to appear within the directory.
- Links are grouped together on web pages based on a criteria such as subject (yahoo)
- Search terms are searched within Yahoo's Pages
  Use subjects and topics that would appear in:
  Yahoo category titles
  Site titles
  Site description (1 sentence)

A subject tree is a subject-oriented index to many thousands of Internet sites. You begin at a high level of subject area, and usually browse your way down to topics and sub-topics.
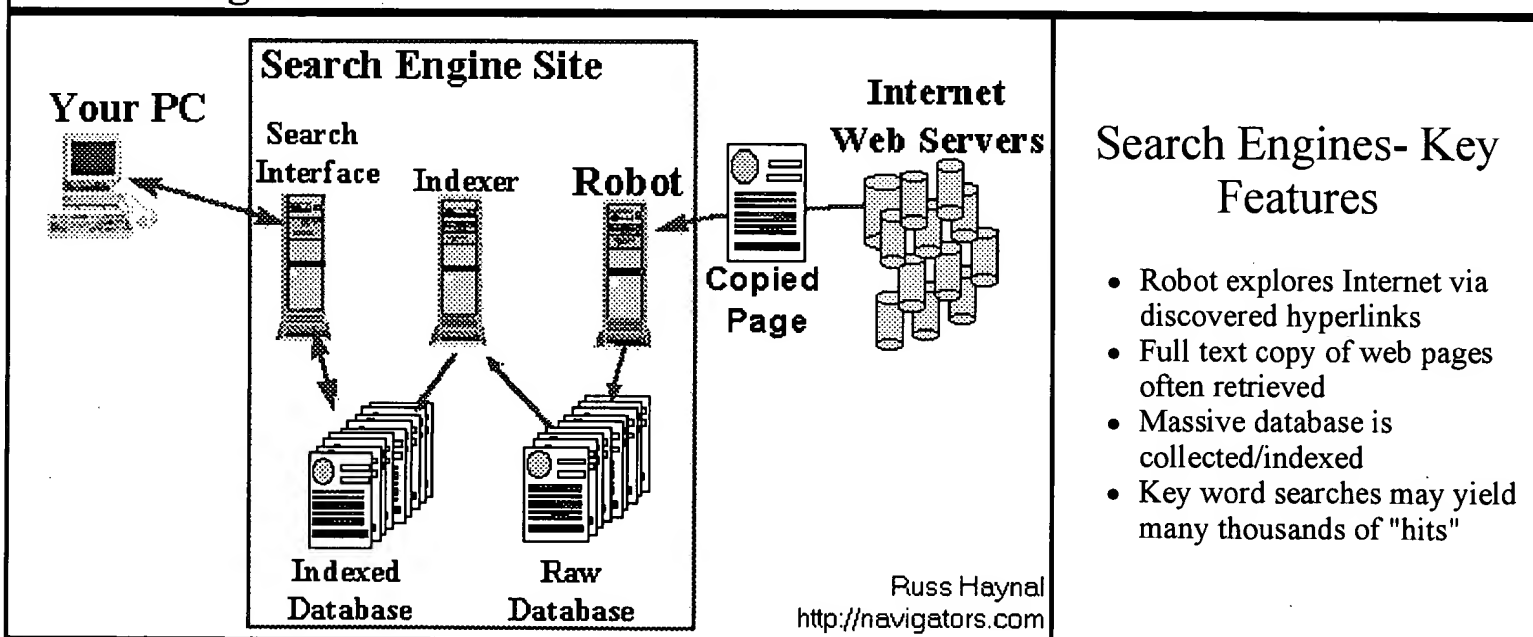
From an initial menu, you must decide "where" your desired subject might be listed. As you select menu items, you will notice that you are going "deeper" down a menu hierarchy. (look at the URL's in the location window) Eventually, you will reach a "bottom page" in the hierarchy. The bottom web page contains the hyperlinks that will take you away from the subject tree web site to the specific web sites containing your desired subject.

A Subject tree may offer a "search" option. Recognize that this is **not** a "search of the Internet", but rather a keyword search of the web pages contained **within** the subject tree. Remember that most subject tree web pages contain **only** the words: "name of subject", "names of subtopics", "names sub-subtopics", "brief description of XYZ website". Therefore keep your search terms at a subject tree at the appropriate level of detail (i.e. search for a subject)

Note that most subject trees are **manually built**, usually from user submissions. If someone creates a web site, and they want to be "found", they should be sure to announce themselves to subject trees such as Yahoo.

One of the oldest and most popular subject trees is Yahoo. It's extensive set of topic pages will often lead you to dozens of online resources related to a specific topic. Main subject categories are indicated in the larger fonts, with popular sub-topics also included in the smaller font. As a popular commercial site, The Yahoo corporation also offers additional online resources such as maps, phone books, and stock quotes.

Search Engines:

Your PC

Search Engine Site

Search Interface   Indexer   Robot

Copied Page

Internet Web Servers

Indexed Database   Raw Database

Russ Haynal
http://navigators.com

Search Engines- Key Features

- Robot explores Internet via discovered hyperlinks
- Full text copy of web pages often retrieved
- Massive database is collected/indexed
- Key word searches may yield many thousands of "hits"

Search engines are usually characterized by very large indexed databases, which contain pointers to millions of URL's. Search engines have two main functions that distinguish them from each other.
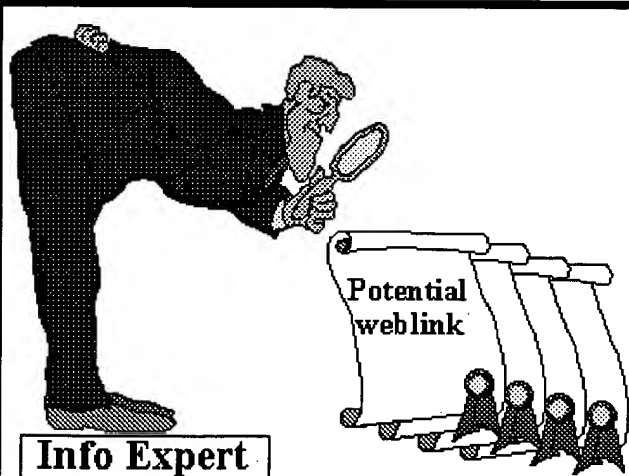
Building the index. Many Search engines develop their vast databases by using a software application to automate the exploration of the Internet. These applications (known as robots, spiders, and crawlers) visit web pages, copies them into a local database, and then explores all the links referenced in the freshly copied page. In this manner, the search engine may eventually discover a high percentage of web pages, so long as they are pointed to from someone else's page. (You can also "invite" the search engine to explore your site) Search engines will NOT copy web pages if they are designated as off-limits by the web site author. See robottxt for more information, and a list of known robots crawlers.

The Search Interface. As the web page information harvested, it is indexed and made available to the Internet through a user interface. Each Search engine may contain different options and parameters that can be used to search its database. In general, you are performing a full keyword search against the entire text from millions of web pages. Search engines also rank the search results in some order based on a scoring criteria. If you use a search engine frequently, it pays to read the "about this search engine" or "advanced Search" options to learn more about how the search engine is interpreting your queries. Note: Most search engines use "or" as the logical operator in a multi-word search. You may want to explore the online help to see how to focus your search queries. Also be careful to avoid using common words which may "dilute" your search results.

Alta Vista is a very powerful Search engine, which offers a wealth of search parameter options. The default interface for Alta Vista is the simple query. If you use Alta Vista on a regular basis, you are strongly encouraged to look at the information contained in their online help. Most Search engines like Alta Vista will respond with a large number of "hits" that match your search string. Fortunately, These search results are "scored" and ranked based on how prominently/frequently the key words appear on the web page.

Each hit includes a hyperlink to the web page and some portion of the web page's text. You should take the time to look at the URL's to help decide which web page might meet your needs before you attempt to access the page. After looking at these top these top ten results, you can then ask for the next batch of ten , and "work" your way through all the suggested web pages. The key to using a search engine is visualizing the web page in your mind, and then composing a specific enough query so that the page you envision will make it into the top 10-20-30 hits.

Virtual Libraries:

Potential weblink

Info Expert

Virtual Libraries- Key Features

- Usually focused on a specific subject area
- Often developed by "experts" in that field

There is an abundance of online resources and a variety of tools to seek out information. Unfortunately, there are few of us who can afford to dedicate "their entire life" to the task of constantly searching the Internet for the latest information in a particular subject. Fortunately, there are other Internet users who **are** able to invest such time into research their (your) topic. Probably the single best way to thoroughly cover a topic, is to discover a Virtual Library for that topic. Virtual Libraries are subject-specific indices that are created (and maintained) by someone who really cares about that topic.

Tip: The best way to discover a "Virtual Library":

1. Go to a good subject tree such as <u>Yahoo</u>
   a. Browse down a couple of layers to the subject of your choice
   b. Usually, there is an "web directories" or "FAQ" listing for each subject area (For example look at <u>this page on Security from Yahoo</u>, and try the "web directories" and "organizations" links)
   c. Go and visit each of the indices for your subject
   d. Make a bookmark of the best ones
2. Always watch for sites labeled: "John Doe's Ultimate guide to widgets" or "The widget meta-site"
3. Ask other users. Within any subject area there, is often 2-3 sites that everyone recognizes as being the best.
4. Try searching for web pages which link to several resources you have already discovered. - For more information on this advanced technique see "<u>surfing upstream</u>"

Discussion: How to choose the right tool

By now you are realizing just how vast the information can be on the Internet. Search engines like Alta vista can quickly provide you with "many" choices. A common question at this point is; "know that we know about Alta Vista, why would we ever use <u>Yahoo</u>?" <u>Yahoo</u> is still quite useful when:

- You only have a general subject in mind such as "web development"
- You don't know enough about the topic.
- You can discover related fields as you browse down the subject tree

For example, to find out about computer security, Yahoo might be your best starting place (computers -> security) Once there, you might learn some of the terminology (i.e. kerberos, pgp, DES) These specific words can eventually be used at a search engines such as Alta Vista.

When startin my search , I ask myself if the topic in question is large enough or smart enough to be announced at Yahoo. For example, If I want **the** FCC web site, I am better off starting at yahoo where I would probably find a listing for the FCC's website. At Altavista, I would gets 300,000 hits of web pages which happen to contain the word fcc.

Search engines are best used when you have very specific information you are seeking. Since Search engines do provide different results from each other, feel free to experiment and try different ones.

Navigate onto the Internet with Russ Haynal
Top-rated Internet Consultant, Speaker, Instructor
Join AT&T, UUNET and others at: <u>navigators.com</u>

Note for my Alumni: Your referrals are **always** appreciated

<u>Home Page</u>	<u>Services / Background</u>
I can be reached at Russweb 'at' navigators.com or 703 -729-1757	
Copyright © 1999 **Information Navigators**	

BPR ⇀⥂⇀ Internet
Making the Connection

This is the On-line companion to my monthly column in

Enterprise Reengineering
The National Publication for BPR

Training With a Net
(From the June 1996 issue)

This month I would like to talk about Internet training. We have all heard negative comments about the Internet like:

- "I can't find what I am looking for."
- "It takes forever to find what I want."
- "Our employees spend their time just fooling around."

Many of these bad experiences can be traced back to inadequate training for Internet users. Just imagine the following scenario:

- You place a bunch of cavemen in running automobiles
- Perhaps you tell them about the gas pedal and the steering wheel
- Do not tell them the rules of the road
- Do not provide any kind of road map

As you walk away from the cars, you can easily visualize the results. One might (incorrectly) conclude that cars and roads are worthless. Yet this is exactly what happens every day on the Internet.

When working with organizations, I explain why Internet training is different and more critical than training for other computer applications.

For example, if you give someone a word processing application and don't train them, the worst that will happen is that they make ugly documents slowly. The word processing user probably won't spend days (or weeks) "playing around" with the font sizes. However, If you give someone Netscape and walk away, you may loose them entirely to the Net.

I have witnessed many disastrous Internet implementations where a company gets a T-1 connection, and Netscape proliferates on everybody's workstation. Thousands of employees then wander aimlessly on the Net in all the wrong directions.

The Internet is very different from most other applications because it includes interactive capabilities. Organizations should recognize that untrained Internet users represent a potential liability to the organization. Users will waste a lot of time, reveal trade secrets and can embarrass the organization in a very public medium.

With one corporate client, I came across a Usenet newsgroup about investing where the company's own employees were commenting on where they thought the company's stock price would be in the future. Clearly, if the Securities Exchange Commission sees this, the company could be in trouble.

Once you decide to invest in Internet training, you want to make sure that the training goes beyond the software mechanics to helping ensure that your employees will use the Internet to support the goals of your organization. Here are several elements that I think are necessary for successful Internet training.

- How to use the software (organizing bookmarks, turning graphics "off", signature files) - this is the simple, obvious part included in all training.
- How to navigate the Internet - reading domain names/URL's.
- subject trees vs search engines
- how to compose a search statement. It may surprise you to discover how many of your employees do not know the difference between "and" and "or" in a keyword search string. (by the way, most search engines assume "or" when multiple words are entered in their search field.)
- Netiquette - Internet users must learn about the culture of Internet if they are to be successful. An organization would never send a manager to Japan to negotiate a contract without first teaching him or her about Japanese culture and language. How many of your employees type their E-mails IN ALL CAPITAL LETTERS, or understand; flaming, IMHO, BTW, and :-) ?
- Access Policy - Employees should be explicitly told/warned what is considered an acceptable use of the Internet at work. Employees should be cautioned about handling proprietary information over the Internet. Employees should also recognize that everything they do goes out onto the Internet with "company.com" associated with it.

  Most corporate firewall/gateways have a transaction log which details every single Internet file/page that goes to every single workstation in the company. Inform employees that this transaction log may be reviewed to see if employees are following the company's Internet access policy. Once employees understand this, 97 percent will be conscious of how they use the Internet and do the "right" thing. What about the remaining 3 percent who continue to spend their entire day looking at playboy.com? This is not an Internet problem, but an employee problem that the Internet can help you document, if necessary.

- The company's Internet class should include a 10-to 20- minute segment on how employees can get a personal Internet connection for home usage. Initially, some of my potential clients question the value of spending the company's time/money teaching employees how to get a personal connection. Employees with a home connection are less tempted to "fool around" on the Internet at work and will become better surfers because of the experience they gain on their own time.

One final consideration for Internet training: The Internet is evolving very quickly. Therefore, training must be thought of as an ongoing process. This investment more than pays for itself when compared to the wasted efforts the average employee expends on ill-fated browsing.

Russ Haynal

Please be sure to visit my Archive of BPR articles.

Home Page	Services / Background
I can be reached at **Russweb@navigators.com** or 703-729-1757	
Copyright © 1997 **Information Navigators**	

Close

How to Compose a Search

_____

The WUTC Search Tool allows you to search for information contained in the web site that may not be readily available through the default views.

Composing Searches:
Type the word(s) and/or phrase(s) you're looking for in the Search Bar. To search for a specific phrase, put the phrase in quotes. When searching for words, type in the words you want using the "logical operators" AND, OR, NOT, etc. Below are some hints for using the various features of the Search Tool.

Logical Operators:

Typing any of these	Finds
NOT not !	documents that do not contain the condition or word following NOT
AND and &	documents that contain both the conditions or words separated by AND
OR or |	documents that contain either of the conditions or words separated by OR
ACCRUE accrue , (comma)	the same documents as OR, but increases the relevance ranking of a document when the document contains more than one matching condition or word
@	documents which contain both of the words that are separated by @, as long as the words are next to each other in the documents. (Do not place spaces between @ and either word.)

Examples:

Typing this text	Finds documents that contain
cat & dog & fish	all three of these words
cat | dog | fish	at least one of these words
cat, dog, fish	at least one of these words. Notes gives higher relevance to documents that contain two or three of these words
dog@fish	dog and fish next to each other
(dog AND cat) OR fish	dog and cat; fish
dog AND (cat OR fish)	dog and cat; dog and fish
dog AND NOT (cat OR fish)	dog, as long as they don't also contain cat or fish

Changing the relevance ranking of certain search words

You can use the Termweight operator to change the relevance ranking of search words, thus increasing or decreasing their importance in the search. Documents in which the search text is most important appear at the top of the list of search results (when you sort the results by relevance).

To use the Termweight operator, type your search text as follows:

- termweight 80 word1 or termweight 20 word2

    where the numbers following "termweight" can be any integer from 0 - 100.

Examples of using the Termweight operator

The table below shows some examples of using the Termweight operator to change the relevance ranking of search

words.

Search text	Results of search
termweight 70 video or termweight 30 audio	Finds documents that contain "video" or "audio" or both. Gives a much higher relevance to "video" than to "audio."
termweight 25 photo or termweight 75 audio or termweight 50 video	Finds documents that contain any of the words "photo," "audio," or "video." Gives the highest relevance to "audio," the next highest relevance to "video," and the least relevance to "photo."

Limiting Search Results:

To change the number of search results you will receive, click on the button to the right of the phrase "No Limit" on the Search Tool. This will pop up a display of your search limit choices
(10, 50, 100, 200, or No Limit). Click on the appropriate number and your search will return the corresponding number of results.

Word Options:

This allows you to set how "picky" your search will be. You have the option of directing the computer to search for the EXACT word(s) that you have entered. Again, this will limit the number of results your search will yield. To set this search parameter, click on the small box next to the phrase, "Find exact word matches only."

Your other option is to direct the computer to search for related words, phrases, or topics as specified in the computer's thesaurus. This will most likely result in a higher yield of search results. To set this parameter, click on the small box next to "Find word variations as defined by thesaurus."

Sort:

Another set of options available to our WebGuests is to tell the computer how to sort the search results. Searching by "Relevance" allows the computer to yield results in its default setting, searching by "Oldest first (by date)" will sort the documents from oldest to latest, and searching by "Newest first (by date) will sort from latest to oldest." To pick your sort setting, simply click on the button next to "Relevance". When the pop-up menu displays your options, click on your sort choice. The computer is now ready to sort its search results according to your wishes.

Date Range:

This option allows you to limit your number of search results by specifying the date a document was created. The computer will look for documents that were created on, before, or after (depending on your choice) the date you specify. To do this, click on the button next to "On", then click on the appropriate choice in the pop-up menu. Then, to the right of that box, type in the date you're looking for. (NOTE: If you do not specify a date, the computer will not use "Date Range" as a search parameter.)

Section: How to Use the WUTC Web Site
Subsection: 6. Searching the Site
Date Composed: 11/05/1997 **Date Modified:** 11/05/1997